# Quantitative Method for Comparing Extracted False Color Images from 2D IR Microspectropic Data to Visual Light Images

Research Thesis

Presented in Partial Fulfillment of the Requirements for Graduation

"with Research Distinction in Chemistry" in the undergraduate colleges of The Ohio State University

by

Nicholas Sunday

The Ohio State University

May 2013

Project Advisor: J. Clay Harris, Departments of Chemistry and Biochemistry

**Abstract**

The purpose of this project was to find a quantitative approach to compare accuracies of visible light and computational images. The computational images that needed comparisons were from the output of a Gaussian mixture model (GMM) program in MATLAB® that was run to separate spectra of cross sections of normal aortic tissue and tissue affected by an aneurysm from background. GMM works by clustering spectral information into related groups for visualization.

By using photo-editing software, it was possible to determine the exact number of pixels by extracting the number of pixels from the computation images and comparing them to the number of pixels from visible light images. From the data that was received, the principle components that were needed for best accuracy were determined. Principal component analysis (PCA) groups information into different orthogonal scores based on decreasing variance in the data. The validity of preprocessing with PCA in order to achieve the best accuracy for the separation of tissue from background was discovered. From the information gathered from this process, it was possible to make improvements in the choosing the most relevant PCAs and verify the improvements quantitatively.

In this project, the different combination of principal components and how they affect GMM clustering were compared. The long-term goal of this project is to use spectra from FT-IR spectroscopy of biological analogs in order to obtain information about changes in the basal structure of tissue so that aneurysms may be detected early in their development.

## Introduction

Abdominal aortic aneurysms (AAA) are becoming a vast burden in many developed countries.[1] Hospital discharge data from 2003 shows that 6,815 patients were discharged with a ruptured AAA with a mean length of stay of 10.7 days and total charges in excess of $639.71 million.[2] In the United States alone, AAAs are the thirteenth leading cause of deaths in people over 65 years of age.[1] The numbers of deaths due to AAAs are rapidly on the rise, partly due to the increasing numbers of elderly individuals in the general population.[3] It is puzzling that AAAs would lead to such a large incidence of death due to the vast methods available for detection and repair. In fact, open surgical procedures for repair only have a mortality rate of approximately 2% at tertiary referral centers.[4] The seemingly inconsistent correlation of deaths vs. mortality rate after treatment results from the deadly nature of aortic aneurysms. AAAs are frequently asymptomatic until rupture, presenting with mild symptoms such as back and/or abdominal pain.[3] The overall mortality rate resulting from ruptured aneurysms is between 65% and 85% with about half resulting from ruptures before a patient reaches a surgical room.[1] Due to the large number of deaths due to missed diagnosis of AAAs, a method of detection that could detect an aneurysm before it leads to weakening of the aorta would be very beneficial.

The majority of AAAs, with 90% occurring below the renal arteries, are associated with atherosclerosis.[5] Atherosclerosis is a chronic inflammatory process with a complex mechanism.[5, 6, 7] Extensive research has been done to understand the complex pathophysiology of atherosclerosis, however, the mechanisms for various

aspects of the disease are still being discovered.[8] The earliest changes in the vessel wall include increases in retained lipoproteins[9, 10, 11] following oxidation in the subendothelial matrix.[9] Risk factors such as elevated plasma cholesterol, hypertriglyceridemia, hypertension, and smoking are found in many patients with AAAs, yet 60% of these cases have plasma cholesterol levels of greater than 240 mg/dL.[5] The National Cholesterol Education Program, Adult Treatment Panels suggests plasma cholesterol levels should be less than 200 mg/dL. This observation is consistent with the fact that the lower abdominal aorta is the site where atherosclerosis first develops, which becomes common by middle age.[8]

A subcommittee appointed by The Society for Vascular Surgery and the International Society for Cardiovascular Surgery has recommended that an AAA can be defined as a permanent 50% or greater increase in infrarenal aortic diameter (IAD) compared with the expected IAD based on age, sex, and other factors.[12] Normal IAD in individuals over the age of 50 are 1.5 cm in women and 1.7 cm in Men.[13]

Most aneurysms are true aneurysms in the respect that they are local dilations of all three layers of the arterial wall: intima, media, and adventitia. A pseudoaneurysm is defined as a collection of flowing blood that is in contact with the arterial lumen but is not contained in the normal vessel wall; it is contained by the adventita or other sounding soft tissue.[1] The formation of a saccular (nondissecting) AAA follows the destruction of the connective tissue in the media, particularly the elastic lamellae.[14]

Elastin and fibrillar collagens are the main components to the mechanical strength of the aorta.[15] The strength that these large biological proteins provide are vital to dealing with high systolic pressure produced by the heart. Along with collagens and elastin, there are thousands of other proteins that make up the extracellular matrix of the aorta; however, changes in the proportion of elastin and collagen have widely been implicated in aneurysm formation, progression, and rupture.[15]

There has been much research done to discover the cause of the degradation of collagen and elastin. Studies have shown that there is an increased local production of metalloproteinases (proteins capable of destruction of elastin and collagen) in patients with AAAs.[16] The vessel wall is then unable to withstand the force of each systolic contraction and destructs.[14] Specific events, most notably changes in collagen and elastin distribution in arterial walls have been associated with the development of aneurysms.[11]

Infrared spectroscopy is a technique that utilizes the fact that molecules absorb specific frequencies that are characteristic of their structure. These differences in absorption can be detected and used to provide a sort of "biochemical fingerprint" of tissue samples. FT-IR spectroscopy has already been utilized in this manner to provide information on myocardial infarctions.[11] In a study by Cheheltani et al, cardiac collagen deposition, a process often seen with myocardial infarction, was visualized by Fourier transformation infrared imaging spectroscopy (FT-IRIS).

In this investigation, clustering techniques were used as a way to separate tissue spectral data from background spectral data of the slides on which the tissue

samples were fixed. Prior to 2D FT-IR analysis, tissue slices were frozen in Optimal Cutting Temperature compound (OCT), and sectioned and mounted on glass slides. Tissue slices were obtained from control mice and experimental genetic knockout mice that were apolipoprotein E-deficient (apoE -/-). Aneurysms were induced by intravenous injection of angiotensin II.[17] Angiotensin II has been linked to cause atherosclerotic lesions and aneurysms in the genetically altered apoE -/- mice.[18]

The output data from the 2D FT-IR analysis included spectra from both the tissue and the slides on which the tissue slices were mounted. Spectra were obtained from 720-4000 $cm^{-1}$ with 6.25 x 6.25 μm spatial resolution and 8 $cm^{-1}$ spectral resolution (Figure 1).[17] In order to help analyze the data, principal component analysis (PCA) was used as a data reduction technique to help analyze the vast output of spectral data; for some tissue slices, spectra obtained were in excess of 60,000.

PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of possibly correlated values into a set of uncorrelated values called principal components. PCA is performed with an eigenvalue decomposition of the covariance matrix of a multivariate dataset. The covariance between two variables in a dataset indicates how they differ relative to each other. If the covariance is negative, then the two variables are inversely related, and if the covariance is positive, then the two variables are directly related. If, however, the variables are independent, the covariance will equal zero. The resulting covariance matrix contains covariance values for all possible combinations of variable pairs of a multivariate dataset.

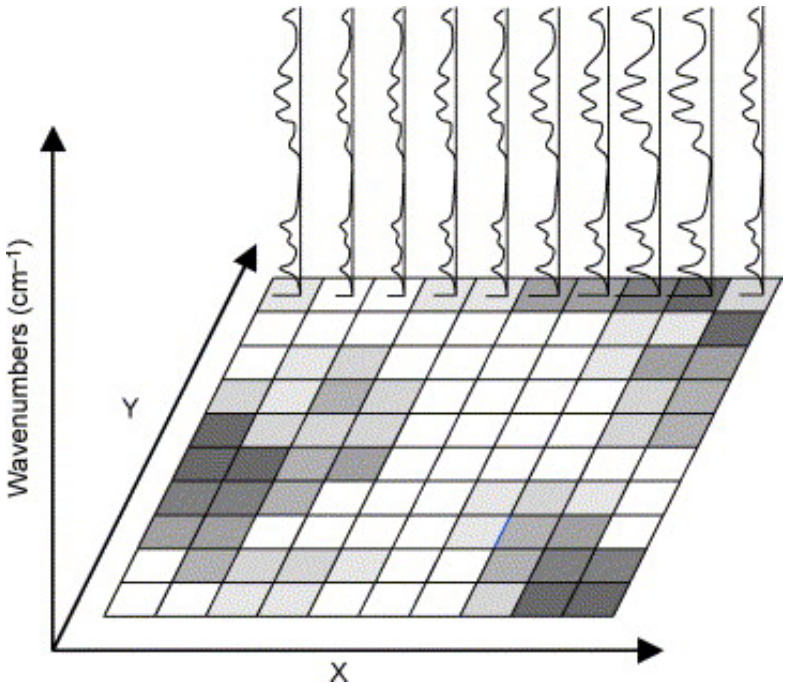


**Figure 1**. Data set format: 720-4000 $cm^{-1}$6.25 x 6.25 μm spatial resolution, 8 $cm^{-1}$ spectral resolution.[17, 19]

Eigenvalue decomposition of the covariance matrix produces sets of eigenvectors and eigenvalues. Each eigenvector is orthogonal and gives information about patterns in the data. Figure 2 illustrates orthogonal eigenvectors. PC1 lies among the greatest variance of data with PC2 orthogonal to it. The eigenvector has a corresponding eigenvalue, which gives information regarding the eigenvectors significance. The eigenvector with the highest eigenvalue is the 1st principal component in that data set, which corresponds for the greatest variance in the data.[20] Through the eigenvalue decomposition procedure, hyperdimensional data can be compressed into a few dimensions so that analysis of data becomes simplified. For purposes of this research, PCA is used for preprocessing to reduce each individual 2D FT-IR spectrum into a single data point.

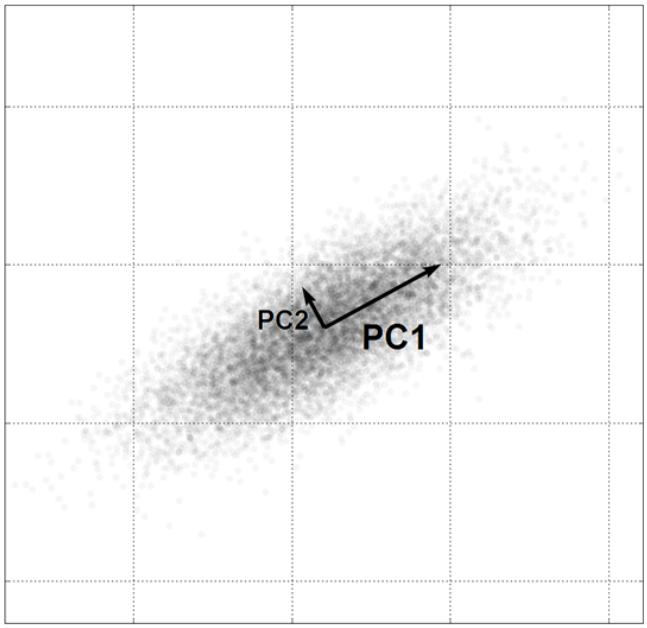


**Figure 2.** Representation of PCA of an arbitrary set of data.[21]

After the spectral sets of data have been reduced with PCA, the clustering algorithm Gaussian mixture model (GMM) was investigated on its ability to cluster spectral information into related groups in order for visualization. GMM (Figure 3) is a very convenient technique in the respect of its unsupervised clustering algorithm as opposed to selecting representative data points, which proves to be lengthy and inaccurate. In a study by Toyama et al., GMM was shown to be effective as a technique for differentiating background and foreground pixels while simultaneously circumventing many common problems associated with background subtraction.[21] GMM operates under an assumption that a given dataset is composed of a finite number of Gaussian distributions. GMM makes initial approximations of the centers for the number of expected Gaussian distributions that have been specified by the operator. These approximate centers are then iteratively refined, as

data points are shifted between clusters until the algorithm converges upon a single value.[16]

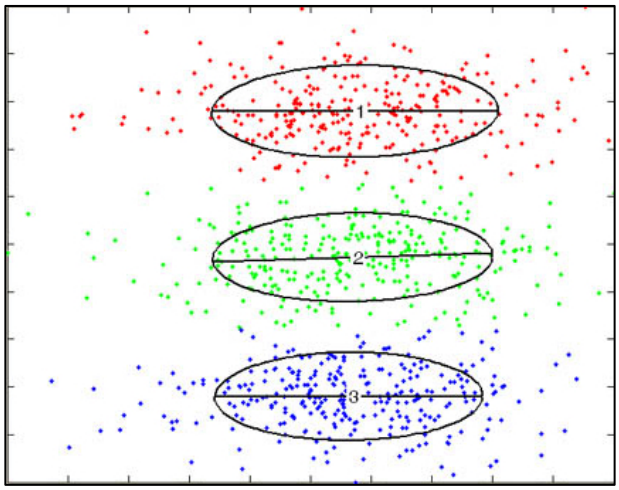


**Figure 3.** Representation of GMM Through an arbitrary set of data.[22] Each circle represents a different clustering assignment.

PCA was found to be a vital preprocessing step to GMM when it comes to successfully clustering images and separating tissue from background. Figures 4 and 5 show the same spectral data set with and without PCA. Figure 4 failed to cluster the spectra into more than one group (as represented by the single solid green color). Figure 5 shows successful clustering with PCA into three distinct clusters. Tables 1 and 2 summarize the percentages of images that had successfully clustered spectra data. The naming convention is c (control) or x (experimental) –ak (above kidney) or bk (below kidney).

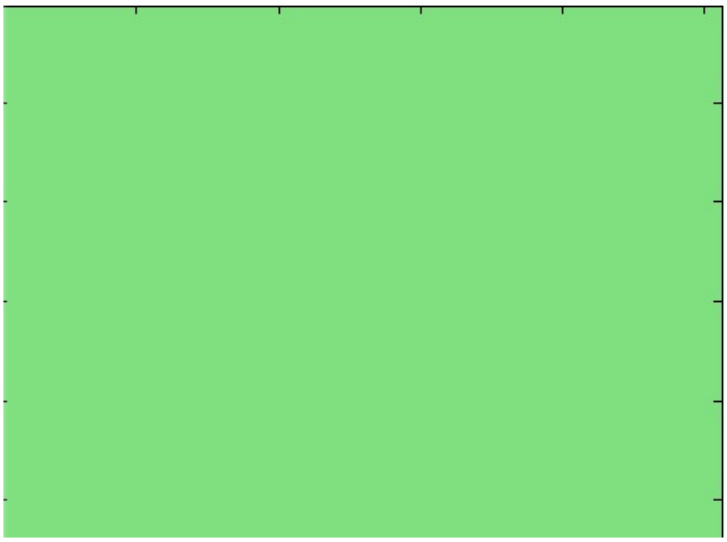

**Figure 4**. Failed clustering of xdatabk1 without PCA.

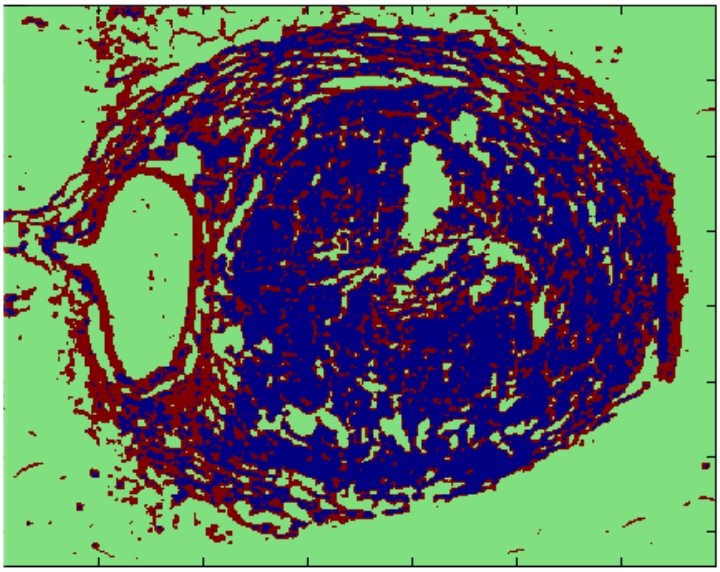

**Figure 5**. Successful clustering of xdataak1 with PCA.

| 2 Clusters | Without PCA | With PCA |
|---|---|---|
| **c-ak** | 83% (22) | 100% (229) |
| **x-ak** | 50% (23) | 100% (239) |

**Table 1**. Results of GMM clustering with and without preprocessing with PCAs 1-4 (assuming 2 clusters). Number of images scaned in parentheses.

| 3 Clusters | Without PCA | With PCA |
|---|---|---|
| **c-ak** | 74% (22) | 100% (229) |
| **x-ak** | 29% (23) | 100% (239) |

**Table 2**. Results of GMM clustering with and without preprocessing with PCA 1-4 (assuming 3 clusters).

*Current imaging techniques*

Ultrasonography (Figure 6 and 7) is the imaging technique of choice for screening and monitoring the rate of AAA enlargement and involvement. Because of its operator dependence, low cost, availability, and speed, it is the principal tool in diagnosing and monitoring AAAs. Ultrasonography has a sensitivity of nearly 100 % in the diagnosis of AAAs and can report aneurysm diameters at an accuracy of 0.3 cm.[3] Oftentimes, the full extent of the aneurysm cannot be visualized fully or it might be completely blocked or obscured due to obesity or bow distinction due to excess gas. As a result, ultrasonography cannot alone fully provide the necessary information required in a preoperative evaluation of an AAA.

Currently, the most widely used imaging technique for evaluating aortic structure and health is through computed tomography (CT). CT scans (Figure 8) are usually preformed as a screening test if an ultrasound images are suboptimal or as a diagnostic test if a ruptured AAA is suspected.

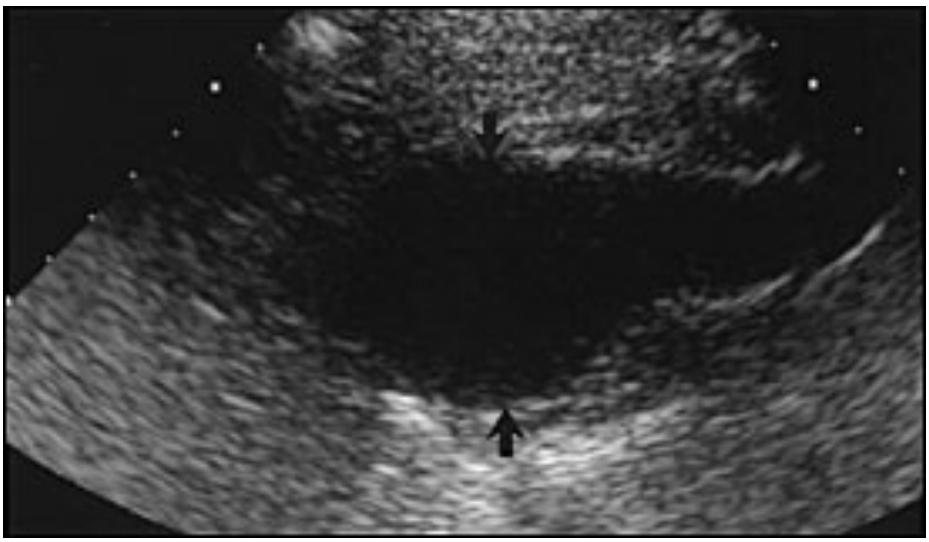

**Figure 6**. Longitudinal ultrasonography of the abdominal aorta with aneurysm.[12]

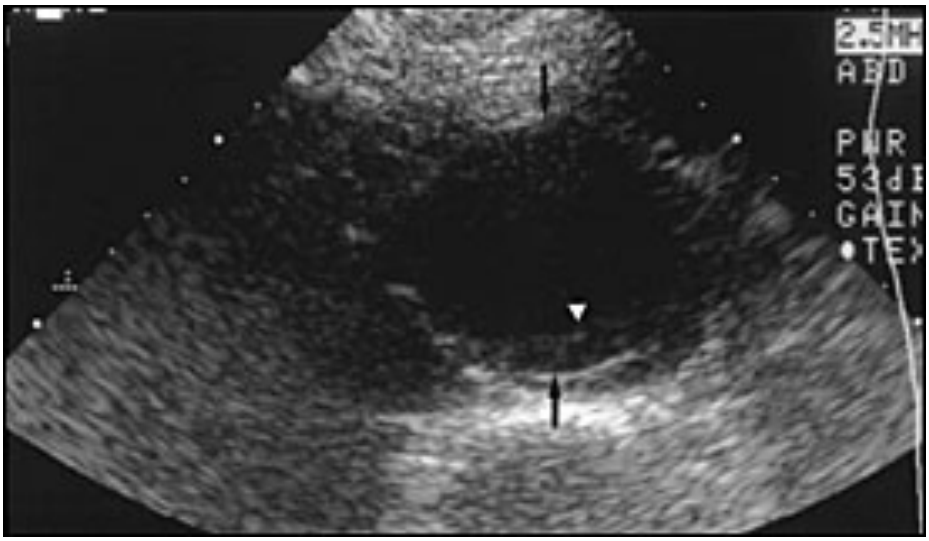


**Figure 7.** Transverse ultrasonography of the abdominal aorta with aneurysm.[12]

CTs are invaluable as a tool for studying aortic aneurysms and various other disorders due to its speed of examination and widespread availability. Unlike ultrasonography, there is no concern regarding reproducibility of results, technical error, or limitations caused by body size or the presence of bowel gas.[3] While CT shows many advantages to ultrasonography, there are limitations to traditional CT. CT images often contain limited information of arterial anatomy and aneurysm

involvement. AAAs usually require additional methods to obtain necessary information such as maps of nearby vessels for planning surgical repair.

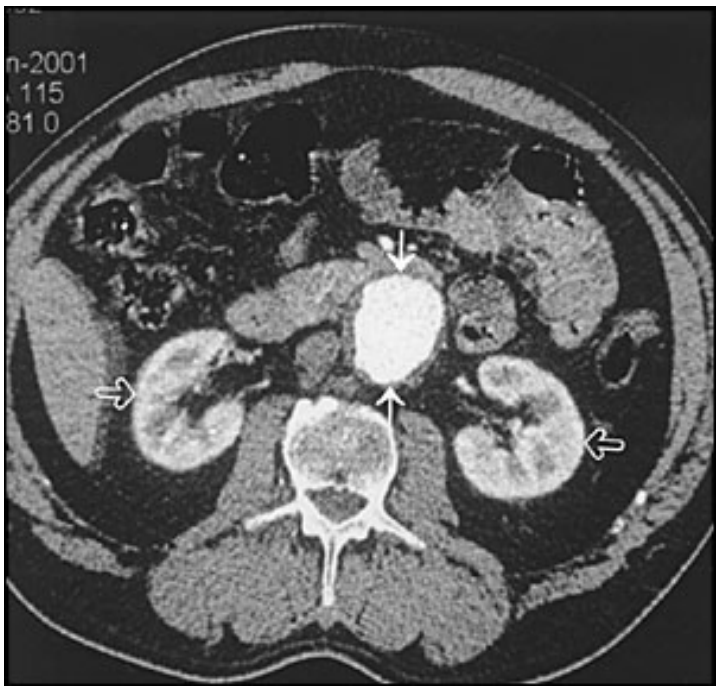

**Figure 8**. CT of an abdominal aorta at the level of the kidneys showing an aneurysm.[12]

Magnetic resonance imaging (MRI) can also be used to determine the extent of AAAs (Figure 9). The absence of iodinated contrast material that is sometimes used in CT scans are what makes this method advantageous. However, MRI is more sensitive to motion that CT. Since patients must remain motionless for a longer period of time for MRI than CT, there is often less detail than CT scans with other organs that might be involved or displaced with the aneurysm. MRI may also be limited to some patients due to the interactions of metals in the patients with the magnet used in the instrument.

**Figure 9**. MRI of an abdominal aortic aneurysm. The lack of information due to other structures is likely due to motion during the lengthy scan.[12]

Aortography now has a limited role in a preoperative evaluation of AAAs due to many drawbacks that prevent its regular use. Aortography tends to underestimate the size of aneurysms (Figure 10) because it only demonstrates a patents lumen and typically there is a circumferential thrombus organized within most aneurysms. Also, its use is limited due to its associated cost, exposure to iodinated contrast, and invasive nature. Aortography is often reserved for special situations, such as when patients have suspected renovascular stenosis, chronic mesenteric ischemia, occlusive iliac disease, juxtarenal or suprarenal AAAs, horseshow kidneys, previous colectomy, and thoracoabdominal extension.[3]

**Figure 10.** Aortography of an AAA. Involvement of other arteries is clearly seen.[12]

**Experimental**

In order to quantitatively determine the accuracy of each cluster assignment (CA) of the successful clustering images, an analysis of the amount of tissue pixels that matched the actual tissue sample needed to be performed. CA and visible light images could be visually overlay in order to get a rough estimate of accuracy, however, a quantitative approach is needed in order to determine which principal components or combinations of principal components are most relevant for imaging purposes. This analysis was done in the software GIMP 2.8.3 (GNU Image Manipulation Program).

The first step for getting an accurate pixel count is to obtain a pixel count of the visible light image of the tissue cross section mounted on the glass slide. Since counting thousands of pixels by hand is a long and laborious task, a technique was developed utilizing the photo editing software GIMP and a pixel counter (Figure 15) in MATLAB® (The MathWorks, Inc., Massachusetts). In order to obtain a pixel count via MATLAB®, a black image of the pixels that needed to be counted must be uploaded into MATLAB®. GIMP was used to turn each image that needed a pixel count into a black image.

To obtain a black representation of the visible light image of the aortic tissue cross section (Figure 11), the visible light image to be studied was opened in GIMP using the "open as layers" option. After the image has loaded into GIMP, the image was zoomed in to 800% in order to see the individual pixels that made up its edges. The intelligent scissors feature was used to then select the outside edge of the tissue. Intelligent scissors is a feature of GIMP which uses a built-in algorithm that connects two arbitrarily selected points by a line that is generated to be between the greatest contrasts of pixels. Once the entire edge of the tissue slice has been selected, the paint bucket (fill whole selection option) was used to color the area inside the selected tissue. Next, in the same manner as the intelligent scissors were used to select for the outside tissue edge, the inside tissue edge was selected. After the inside tissue slice has been selected, the area inside the interior tissue edge was colored back to white.

After the black image representation of the light micrograph image has been generated, the CA image must be overlaid with the original light image (Figure 12).

This is done by opening both images in GIMP under the “open as layers” option, then changing the opacity of the top image to 50% on the layers toolbar in order to view both images overlaid.

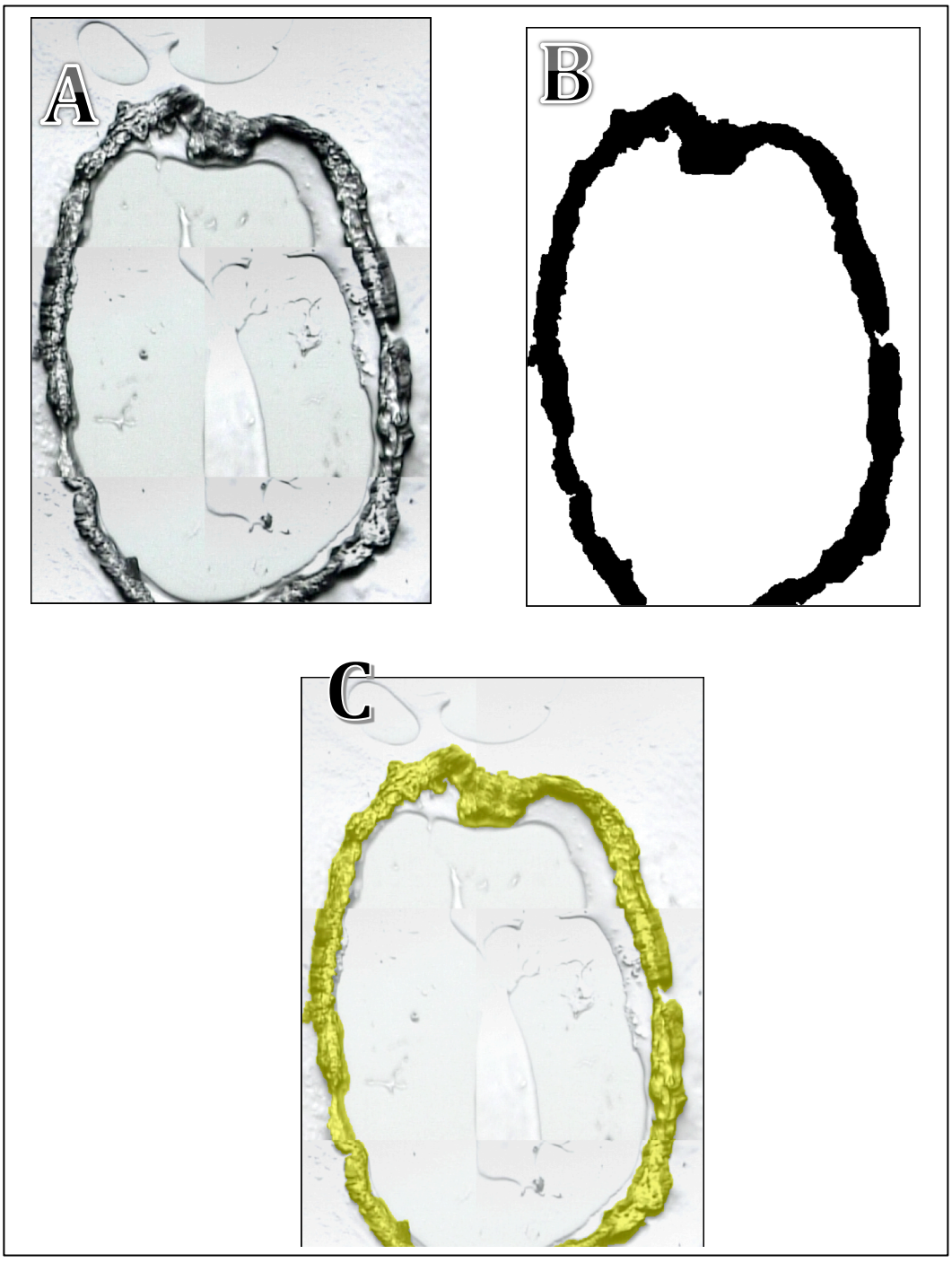


**Figure 11**. (A) Original tissue light micrograph, (B) Extracted tissue pixels, (C) Overlaid light micrograph and extracted tissue pixels (yellow).

The images then need to be resized so that they overlay correctly. Resizing is performed by selecting the “scale tool” from the main toolbar and changing the size

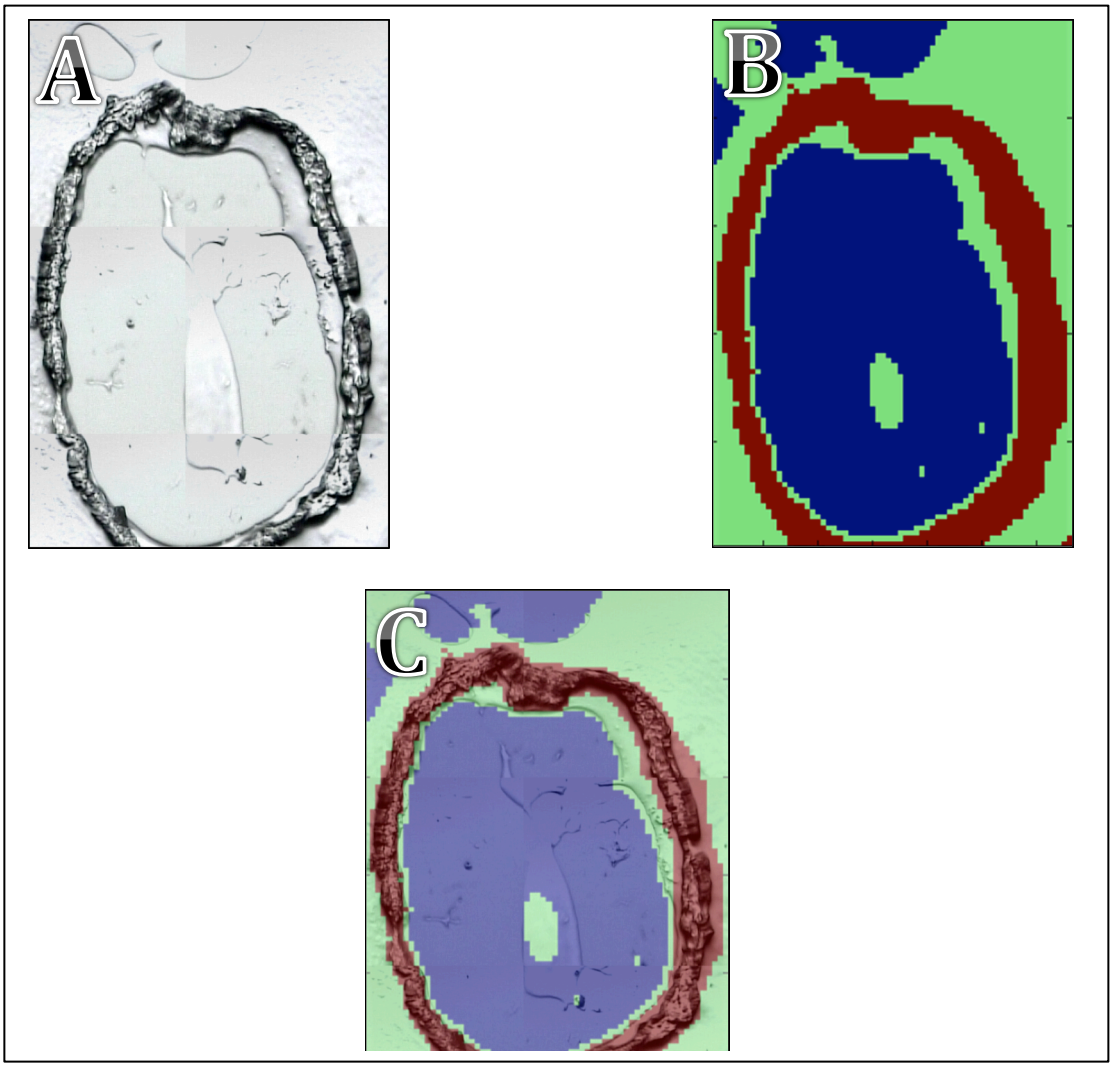


**Figure 12**. (A) Original tissue light micrograph, (B) PC1+PC2 of cdataak1, (C) Overlaid light micrograph and PC1.

and location of the top image until is overlaid correctly with the bottom image.

After the image has been resized, the CA that corresponds to the tissue (Figure 13a), in this case, red, must be isolated and given a pixel count. This is done in the same manner as described earlier. The last step is to generate black images that correspond to the parts of the CA that overestimated regions (Figure 14). This is, once again, done with the same procedure as illustrated earlier. Once all these black images corresponding to the CA and tissue slices have been generated, they can then be uploaded into MATLAB®'s pixel counter (Figure 15). The pixel counter only counts black pixels, which is why the images that needed pixel counts needed to have back representations made.

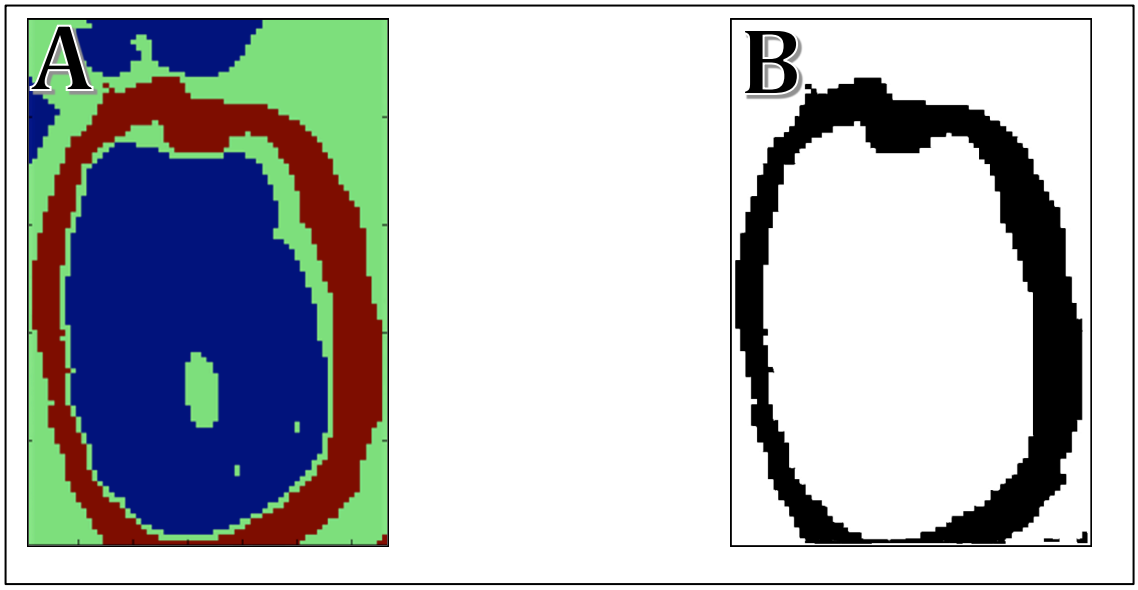


**Figure 13.** (A) PC1+PC2 of cdataak1 (B) Extracted tissue cluster

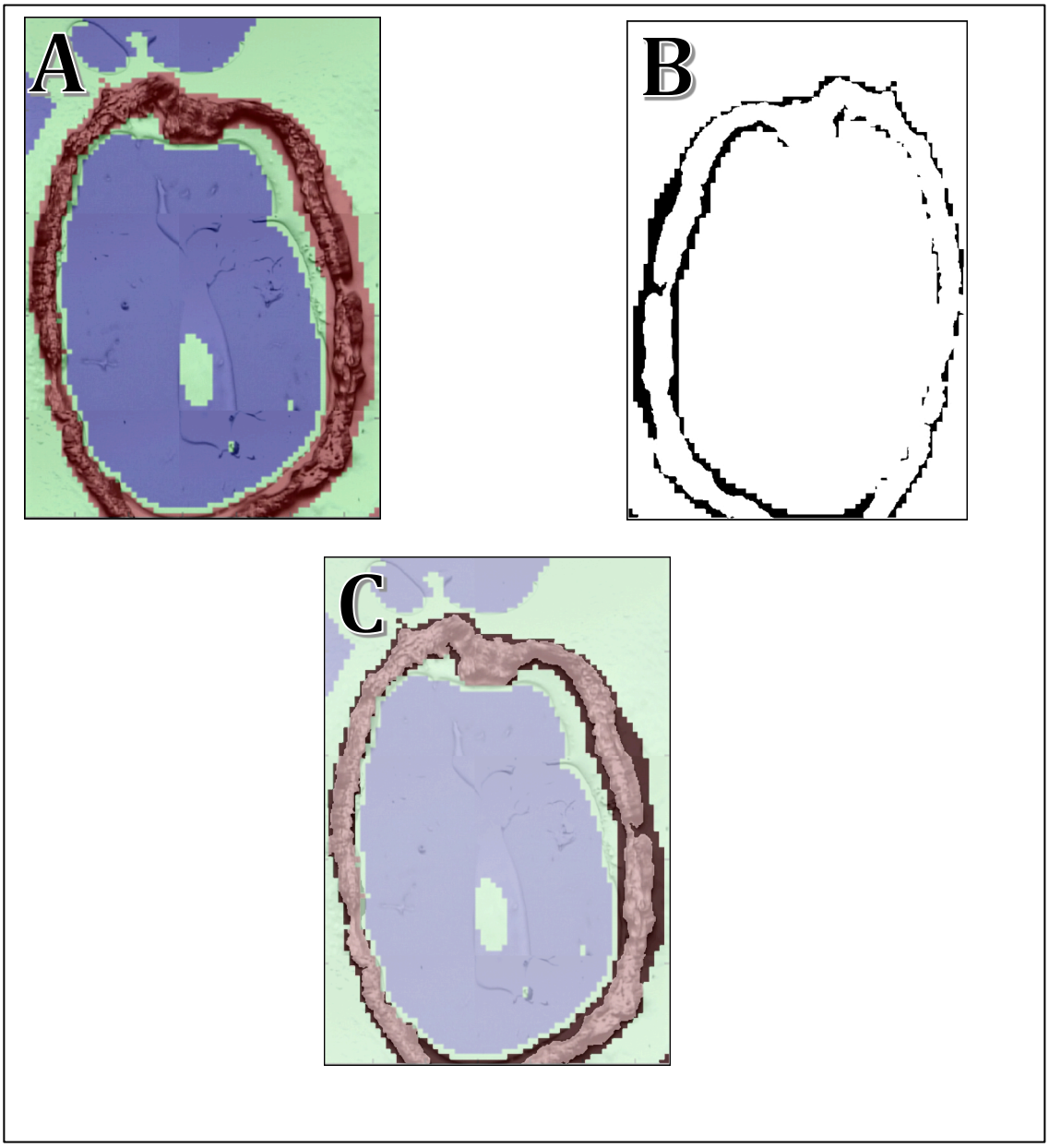


**Figure 14**.(A) PC1+PC2 of cdataak1overlaid with original light micrograph (B) Extracted overestimation of pixels, (C) Overlay of PC1+PC2, original tissue light micrograph, and extracted overestimation of pixels.

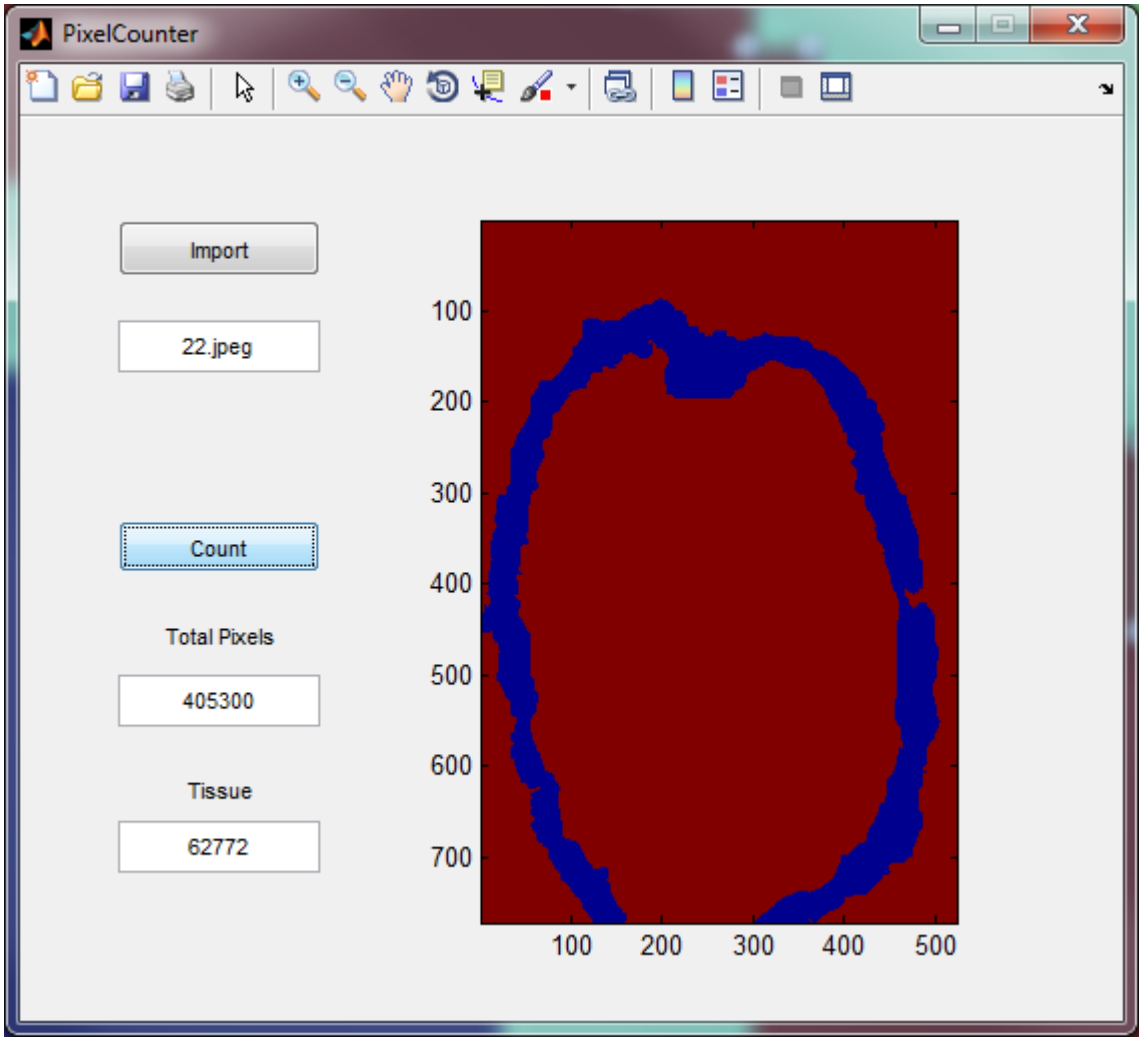


**Figure 15.** Pixel counter that was used for pixel counts.

## Results and Conclusions

After pixel counts had been generated, they were compared in order to obtain relative numbers on the accuracy of each CA. There were a variety of PCs used in the study (PC1, PC2, PC3) and combination of PCs (PC1+2, PC1+3, etc). Table 3 summarizes the percentages of how PCs fit the visible light images and the relative accuracy of PC1 and PC1+PC2.

| Column1 | Original tissue | Original guessed | Pixel overestimation | Percent Correct | Percent Error |
|---|---|---|---|---|---|
| PC1 | 62,772 Pixels | 88,257 Pixels | 25,968 Pixels | 71.12 % | 29.42 % |
| PC1+PC2 | 62,772 Pixels | 89,064 Pixels | 26,768 Pixels | 70.48 % | 30.05 % |

**Table 3.** Comparison of PC1 and PC1+PC2

The percentage of pixels that were correctly grouped by the clustering analysis was calculated by dividing the number of pixels in the original tissue light micrograph by the pixels clustered using the clustering analysis and multiplying by 100. Percent error was calculated by dividing the original number of pixels guessed by the clustering analysis by the number of pixels overestimated and multiplying by 100. Ideally, the percent correct and percent error should add up to 100%, however, because of systemic error, it was slightly skewed.

Because PC1 accounts for the most variety of data, it was hypothesized that PC1 would fit the visible light image the most. This result was verified through the completed analysis. PC1 has the highest percent correct of pixel matching and the lowest percent error. Overlaying PC1 and PC1+PC2 shows that the images are very similar except for a few areas where PC1+PC2 spreads farther than PC1 (Figure 17). Higher orders of PCs such as PC2 were not included in the analysis of data because accuracy dropped of significantly after PC1 and combinations of PC1. Figure 16 shows the relative accuracy of PC2. In comparison to PC1+PC2, PC2 clustering accuracy was very poor.

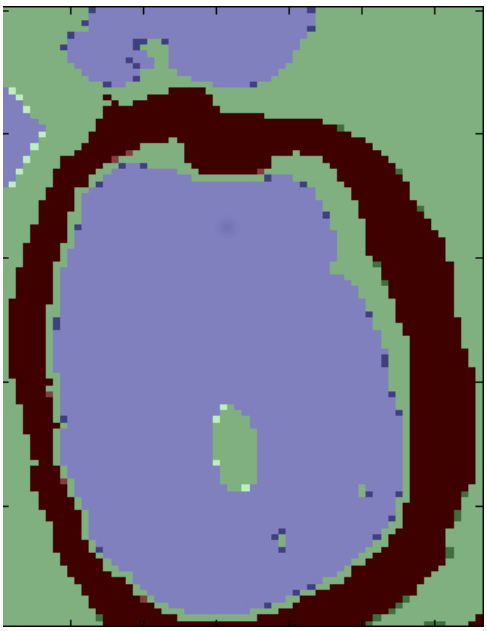

**Figure 16.** PC1 overlaid with PC1+PC2. Bright spots correspond to where PC1+PC2 outreaches PC1.

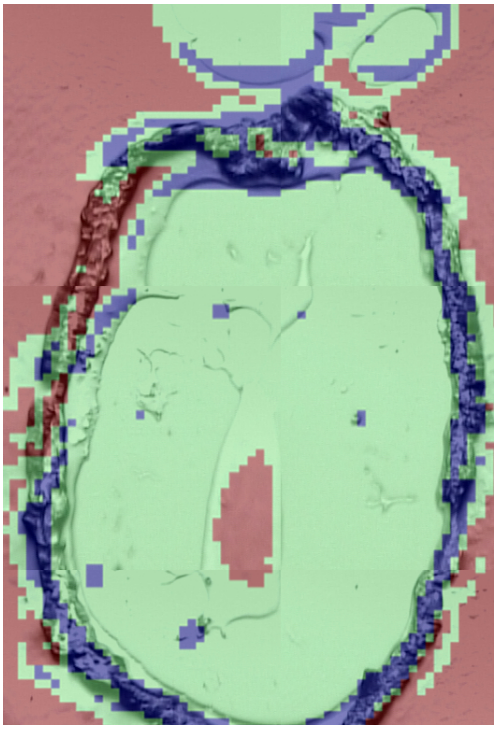

**Figure 17**. PC2 overlaid with the original tissue light micrograph.

From the analysis of pixels, it was confirmed that PC1 was the best representation of the data in terms of clustering spectral data. Although using spectral information could be developed into a method for detecting aneurysm formation in early stages, it will never replace modern scanning methods available. While spectral information gives clues of a developing aneurysm, techniques such as MRI and CT are needed in order to give surgeons vital information needed for surgical repair. However, spectral data and modern scanning methods when used together could potentially reduce death rates for aneurysms and ease the burden that they place on families and the medical community.

## References


1. Sakalihasan, N., Limet, R., & Defawe, O. D. Abdominal Aortic Aneurysm. *The Lancet* **2005***, 365*, 1577-1589.

2. Agency for Healthcare Research and Quality. *Hospital Cost and Utilization Project: Nationwide inpatient sample data set*. Rockville, MD: Agency for Healthcare Research and Quality; **2006**.

3. Upchurch, G. R.; Schaub, T. A. Abdominal aortic aneurysm. *American Family Physician* **2006**, *73*, 1198–204.

4. Hertzer, N. R. Current Status of Endovascular Repair of Infrarenal Abdominal Aortic Aneurysms in the Context of 50 Years of Conventional Repair. (H. Kuivaniemi, M. D. Tilson, & G. R. Upchurch, Eds.) *The Abdominal Aortic Aneurysm: Genetics, Pathophysiology, and molecular biology* **2006***, 1085*, 175-186.

5. Mitchell, M. E.; Siday, A. N. Atherosclerosis is an inflammatory disease. *Semin. Vasc. Surg.* **1998**. *11* (3), 134-41

6. Urbas, A.; Manning, M. W.; Daugherty, A.; Cassis, L. A.; Lodder, R. A. Near-Infrared Spectrometry of Abdominal Aortic Aneurysm in the ApoE$^{-/-}$ Mouse. *Analytical Chemistry* **2003**, *75*, 3650–3655.

7. Nordestgaard, B. G.; Nielsen, L. B. Atherosclerosis and arterial influx of lipoproteins. *Curr. Opin. Lipidol*. **1994**, *5* (4), 252–7.

8. Dobrin, P.; Mrkvicka, R. Failure of elastin or collagen as possible critical connective tissue alterations underlying aneurysmal dilation. *Cardiovascular Surg.* **1994**, *2*, 484–488.

9. Hamilton, C. A. Low-density lipoprotein and oxidised low-density lipoprotein: their role in the development of atherosclerosis. *Pharmacol. Ther.* **1997**, *74* (1), 55–72.

10. Williams, K. J.; Tabas, I. The response-to-rentention hypothesis of atherogensis reinforced. *Curr. Opin. Lipidol.* **1998**, *9* (5), 471–4.

11. Cheheltani, R.; Rosano, J. M.; Wang, B.; Sabri, A. K.; Pleshko, N.; Kiani, M. F. Fourier transform infrared spectroscopic imaging of cardiac tissue to detect collagen deposition after myocardial infarction. *Journal of Biomedical Optics* **2012**, *17*, 056014–1.

12. Sparks, A. R., Johnson, P. L., & Meyer, M. C.. Imaging of Abdominal Aortic Aneurysms. *American Family Physician* **2002***, 65* (8), 1565-1570.

13. Reed D, Reed C, Stemmermann G, Hayashi T. Are aortic aneurysms caused by atherosclerosis? *Circulation*. **1992**; 85:205–211.

14. Ross, R. Arthrosclerosis – An Inflammatory Disease. *N. Eng. J. Med.* **1999**, 340 (2), 115-26

15. Ross, R. *Am. Heart J.* The pathophysiology of atherosclerosis. **1991**, 138, S419-20.

16. Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B. Wallflower: Principles and practice of background maintenance. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* **1999**; Vol. 1, pp. 255–261.

17. Harris, J. C. New Bioinformatic Techniques for the Analysis of Large Datasets. Ph.D. Dissertation, University of Kentucky: Lexington, KY, 2007.

18. Daughtery, A.; Manning, M.; Cassis, L**.** Angiotensin II promotes atherosclerotic lesions and aneurysms in apolipoprotein E-deficient mice. *The Journal of clinical investigation*. **2000,** 105(11), 1605-12

19. Mills ENC, Parker ML, Wellner N, Toole G, Feeney K, Shewry PR. "Chemical imaging: the distribution of ions and molecules in developing and mature wheat." J. Cereal Science. 2005; **41**: 193-201.

20. L. Smith, A Tutorial on Principal Components Analysis, www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf, **2002**.

21. FranzDale, Ben. *Gaussian Scatter PCA*. 2009 <http://en.wikipedia.org/wiki/File:GaussianScatterPCA.png>.

22. Ververidis, Dimitrios. Gaussian Mixture Modeling GUI (GMM DEMO). *Mathworks*. 2009 <http://www.mathworks.com/matlabcentral/fx_files/23848/3/Verver_DEMO_GMM_Version7p1p4out.jpg >.